

## The Evaluation of the Regional Logistics Development in Shandong Province Based on Biclustering Analysis

Zhang Lei<sup>1</sup>, Ma Li<sup>2</sup> and Zhao Peixin<sup>3\*</sup>

<sup>1</sup>Computer Science Department, Beijing University of Civil Engineering, and Architecture, Beijing, China

<sup>2,3</sup>School of Management, Shandong University, Jinan, China

<sup>1</sup>lei.zhang@bucea.edu.cn, <sup>2</sup>mali09192421@163.com, <sup>3</sup>pxzhao@sdu.edu.cn

### Abstract

*Currently, the uneven distribution of logistics resources, the imbalance between supply and demand, and the unreasonable structure in the development of regional economy are important topics for governments and researchers. Innovatively, this paper uses biclustering analysis, which is a widely used method in bioinformatics, to evaluate regional logistics development levels. This paper deeply analyzes the similarity of similar regions under certain indicators from logistics demand as well as basic supply capacity based on the dimensions of logistics indexes and the actual circumstances of cities, and puts forward relevant suggestions through analyzing problems of logistic development of 17 cities in Shandong province. The results indicate that biclustering analysis can be very useful in providing reference for policy making on regional logistics development.*

**Keywords:** logistics; evaluation; biclustering analysis

### 1. Introduction

In the 21st century, the economic globalization has developed rapidly and the post-industrial era has arrived. The tertiary industry, especially the service industry, began to break through the national boundaries and developed dramatically around the world. In major developed countries, economic structure has transformed from "industrial economy" to "service economy". China has been placing an emphasis on the transformation and upgrading of industrial structure in order to promote economic sustainable and harmonious development. In the economic globalization environment, the commercial trade, cargo circulation and information communication among countries and regions is becoming more frequent. Then modern logistic emerged and developed rapidly, it plays an increasingly significant role in social and economic development of each region. The level of logistics development has started to become an important measure of economic growth and the degree of modernization of a nation or a region. Therefore, the logistics industry is also known as the "accelerator of economic operation and growth."

Since the "Second Five Year Plan" was introduced, the economic and social development of Shandong Province is entering a crucial period of economic and industrial restructuring. In this background, the logistics industry has become a emerging pillar industries of the service industry in

Shandong Province. The added value growth rate of logistics industry in Shandong province is accelerating during this period. The logistics industry has been constantly expanded, the logistics infrastructure has been improved, and the economic connection between logistics industry and Shandong Province has become ever closer. Although the development of logistics industry is of great importance for promoting economic

---

\*Corresponding Author

transformation and strategical adjustment of industrial structure, there are many issues including inconsistency of logistics development among different cities, the imbalance between supply and demand, and the unreasonable structure. How to make a reasonable assessment of urban logistics supply and demand levels, and how to take appropriate differentiation strategies based on the different characteristics, categories and constraints for cities with different levels of logistics development are crucial to get the logistics industry off the ground.

Cluster analysis, a method for statistical data analysis, can cluster data effectively by various algorithms that differ significantly in their notion of what constitutes a cluster and how to efficiently find them. Computer aided is also necessary for an efficient clustering process. Cluster analysis method continues to be popular among researchers and has been applied widely in many different fields. Scholars from mathematics, sociology, statistics, science and many others fields modified and improved this method from different perspectives, and have successfully applied related methods in economic management, marketing, social, cultural, biological information and other fields

The traditional ideas of clustering is to cluster rows or columns of a data matrix, and results are global information that contain all rows or columns. When the number of data dimensions and attribute index continue to improve, global information obtained through traditional clustering method may fall short of practical value, and many higher-value local structures that implicit in high-dimensional data are hard to be found. In order to overcome the limitations of traditional clustering algorithms in mining local information in high-dimensional data, Hartigan[1] first proposed the concept of biclustering in 1992. Currently the common biclustering algorithms are mainly Cheng and Church [2], SAMBA[3], CTWC[4], FLOC[5], QUBIC[6], *etc.*, Cheng and Church proposed specific biclustering algorithm and applied it in the research of biological gene expression. The method uses greedy iterative search strategy to find biclustering through the way of moving rows and columns, the authors were interested in finding biclusters with a small mean squared residue called MSR, which is a measure of bicluster homogeneity. They first find sub-matrix to meet the condition that  $MSR \leq \delta$ , while  $\delta$  is the maximum clicket value of MSR, and then maximize its capacity to obtain biclusters. CC algorithm is more efficient in finding a specific number of biclusters, but it is unable to find biclusters with overlapping structure.

SAMBA(Statistical-Algorithm Method for Bicluster Analysis) integratedly uses statistical theory and graph theory, and employes a brute-force solution to identify better biclusters from data matrixes. This method applies a different search strategy: a model for the input data matrix is built in order to establish a totally balanced dipartite graph, then the problem is transferred to find out the dense subgraph in bipartite graph. Since this algorithm adopts brute-force policy, it shares a higher time complexity. Getz *et. al.*, came up with Coupled two-way clustering analysis in 2000, adopting the method of iteration merged rows and columns to find biclusters. The main idea of this method is to identify subsets of rows and columns, and then uses hierarchical clustering algorithm to produce stable clusters by iterations. This algorithm can avoid producing too many meaningless clusters through only allowing stable clusters identified in the last iteration to enter the next iteration.

CTWC method is more flexible on algorithms and easier to be achieved since it can be executed by any conventional clustering algorithm. But the shortcoming of this algorithm is the number of biclusters produced may be large, so that the result is not intuitive enough.

In this article, biclustering analysis method was innovatively applied onto the evaluation of regional logistics development to explore key influence factors on logistics economy. Instead of generically admitting the existence of similarities among all indicators, we revealed the distributions of different indicators for different cities through the studies of city regional clusters. Therefore we deeply realized the intrinsic principle

and trend within the revolutionary development of modern logistics, providing a thorough decision references for a more pinpointed strategies for regional logistic development.

## 2. Source of Data and Preprocessing

### 2.1. Source of Data

To better illustrate the problem, logistics demands and development data from 2003 to 2014 in Shandong province are chosen. From logistics scale and infrastructure perspectives, 13 demand indicators and 10 supply indicators are selected to construct the initial demand matrix and supply matrix in 17 by 156 and 10 by 120, respectively. Because the three-dimensional data of regional logistics development incorporates both indicating and geographic characteristics, dimension reduction is needed for cluster analysis.

### 2.2. Data Preprocessing

Currently, there are numerous evaluation methods of economic development data. Based on the characteristics of the sampling data in this article, two main categories of methods are considered: the first category refers to static evaluation method based on cross-section data, including major composition analysis, factor analysis, level analysis, AHP, entropy weighing analysis. Static evaluation method directly analyzes the characteristics of research targets, but it cannot reflect the influence from time domain. In comparison, the dynamic comprehensive evaluation method is an extension to static evaluation method with the introduction of time factor, thus has been broadly applied in economic evaluations including evaluation-comprehension, comprehension-evaluation, dynamic indicator, and ideal matrix methods. In this article, the time factor discount method is applied to data in time domain to generate comprehensive evaluation data for the next-step cluster analysis.

Weighting time factor is the key problem in evaluating time factor discount method. The authors applied a time discount factor method proposed to solve the weight. The introduction of time factor  $\lambda$  enables a heavier weighing in more recent data to accurately reflect the time influences.  $\lambda$  reflects different weight in time-domain. When  $\lambda$  is towards 1, earlier-stage data is focused while when  $\lambda$  is towards 0, later-stage data is focused. When  $\lambda$  is 0.5, data from different stages are treated equally.

Time weigh  $\omega_t, t \in (1, n)$  satisfies the following conditions

$$\begin{cases} \omega_{t+1} \geq \omega_t \geq 0 \\ \sum_{t=1}^n \omega_t = 1 \end{cases} \quad (2.1)$$

$$\omega_t = \gamma \exp(\lambda t)$$

$\gamma$  is the adjustment coefficient, which could achieve normalized weight under certain conditions of  $\lambda$ . In the general applied research, the value of  $\lambda$  is  $(1/2t)$  and  $t$  as time series number. The paper take the related logistics data from 2003 to 2014 in Shandong province into study, which has twelve time nodes, so we set  $\lambda$  equal to 0.042 approximately and get  $\gamma$  equal to 0.0629 by (2.1). When we take  $\lambda$  and  $\gamma$  into the formula above, we could get the weight vector by MATLAB  $\omega = \{\omega_1, \omega_2, \dots, \omega_{12}\} = \{0.0658, 0.0689, 0.0721, 0.0754, 0.0790, 0.0826, 0.0865, 0.0905, 0.0947, 0.0991, 0.1037, 0.1085\}$ .

We can reduce the twelve dimensional data into one by comprehensive integration of the seventeen cities' time series data of the logistics development in Shandong province, which is beneficial to evaluate the level of the overall development in the low

dimensional level. For the index  $g_j (j = 1, 2, \dots, m)$ , we define the synthesis evaluate function of object  $i$  under the condition of index  $j$  :

$$y_i(g_j) = \sum_{t=1}^N \omega_t x_{ijt}, i = 1, 2, \dots, n \quad (2.2)$$

$x_{ijt}$  is the performance value of the object  $i$  in the year of  $t$  under conditions of the index  $j$ . For each  $g_j$ , a data matrix of  $12 * 17$  is formed  $U_j$ .

$$U_j = \begin{bmatrix} x_{1j1} & x_{1j2} & \dots & x_{1jt} \\ x_{2j1} & x_{2j2} & \dots & x_{2jt} \\ \vdots & \vdots & & \vdots \\ x_{nj1} & x_{nj2} & \dots & x_{njt} \end{bmatrix} \quad (n = 17, t = 12)$$

Take the comprehensive evaluation expression vector of the seventeen cities' relative indexes  $g_j$  as  $H_j = \{x_{1j}, x_{2j}, \dots, x_{nj}\}^T$ .

$$H_j = U_j \omega^T = \begin{bmatrix} x_{1j1} & x_{1j2} & \dots & x_{1jt} \\ x_{2j1} & x_{2j2} & \dots & x_{2jt} \\ \vdots & \vdots & & \vdots \\ x_{nj1} & x_{nj2} & \dots & x_{njt} \end{bmatrix} \begin{bmatrix} \omega_1 \\ \omega_2 \\ \vdots \\ \omega_n \end{bmatrix}, (n = 17) \quad (2.3)$$

After the data is processed, the data set is reduced to the two dimensions of the city indexes in the factor of time. In order to eliminate the influence of the index dimension and make the data comparable, we standardize the data by using SPSS.

### 2.3. Clustering Tools and Methods

In this paper, we use the g-CLUTO (Graphical Clustering Toolkit) which is developed by Matt Rasmussen and others from the University of Minnesota. The software is used to make biclustering for low and high dimensional data sets, and the cluster analysis is carried out. g-CLUTO is the graphical front-end of CLUTO data clustering library, which is used to provide users with a friendly graphical interface. Besides, g-CLUTO also provides several methods of visualization interactive clustering results. There are four clustering methods embedded in the software, which are Repeated Bisection (RB), Direct, Agglomerative and Graph.

We make data into cluster by using g-CLUTO, and get clustering report and visualization solutions including visual matrix and hills. The value of the data matrix is showed by colors in the visual matrix. The data close to zero is in white color, positive value in red and getting deeper in higher numbers, negative value in green and getting deeper in smaller numbers. Visual hills are the hills clustered by Gaussian curve, in which each hill represents a category, describing its clustering information by the volume, height and color of the peak. The volume of the peak is in direct proportion to the number of the objects, that bigger volume represents more objects; The height is in direct proportion to the similarity in the category, and the steeper hills represent higher similarity. There are five kinds of colors of the hills, red, yellow, green, light blue and dark blue. Blue represents the standard deviation in high similarity, while red indicates the standard deviation in low similarity. The objects have lower dissimilarity and higher similarity when the color is closed to dark blue. We divide seventeen cities in Shandong province into three or four categories for comparison purpose. According to the result of cluster by

using clustering methods, similar function, discrimination function and other parameters in many combination experiments, we choose the aiming objects, which are of high average ISim, low average ISdev, low average ESim and low average ESdev.

According to the comparison results, the data matrix is clustered by using RB algorithm, cosine similar function and I2 discrimination function for Biclustering. The optimal parameter combinations of all kinds of methods are shown in Table 2-1.

**Table 2-1. Comparison of Optimal Parameter Combination Settings**

Number	Alg	similar function	discrimination function	Graph Model	Nearest Neighbors	AVG ISim	AVG ISdev	AVG ESim	AVG ESdev
4	RB	C	I2	A-D	4	<b>0.785</b>	<b>0.065</b>	<b>0.210</b>	0.075
4	A	C	SL	A-D	4	0.776	0.087	0.322	0.081
4	G	CC	I2	A-D	5	0.593	0.083	0.178	<b>0.066</b>

### 3. Analysis on the Logistics Development Level Based on Biclustering

#### 3.1. Analysis on the Logistics Requirement Development Level Based on Biclustering

The ISims of the four clusters in Table 2-1, are all greater than 0.6, and the values of ESim, Isdev and Esdev are all on the small side, which indicates that the clustering effect is good. Detailed analysis is shown in Table 3-1, Figure 3-1, Figure 3-2, and subsequent classification analysis.

**Table 3-1. Comparison of Urban Clustering Results**

Cluster 0 Size 5 Isim 0.683 Isdev 0.112 Esim 0.490 Esdev 0.080								
Jinan, Qingdao, Zibo, Yantai, Weifang								
Descriptive	Regional passenger turnover quantity	14.0 %	Regional passenger traffic volume	13.4 %	Total retail sales of consumer goods	10.3 %	The volume of investment in fixed assets	9.9%
Discriminating	regional passenger traffic volume	13.1 %	regional passenger traffic volume	12.3 %	Added value of Secondary industry	10.6 %	GDP	10.5 %
Cluster 1 Size 6 Isim 0.809 Isdev 0.052 Esim 0.224 Esdev 0.035								
Laiwu, Rizhao, Dezhou, Dongying, Bingzhou, Weihai								
Descriptive	Population density	15.9 %	Regional total goods amount	14.8 %	Business total of postal service	9.5%	Total retail sales of consumer goods	9.1%
Discriminating	Population density	28.2 %	Added value of primary industry	15.2 %	Regional total goods amount	15.1 %	Business total of postal service	7.6%
Cluster 2 Size 2 Isim 0.738 Isdev 0.000 Esim 0.178 Esdev 0.05								
Linyi, Jining								
Descriptive	Added value of primary industry	29.7 %	Regional freight kilometers	28.7 %	Regional total goods amount	28.0 %	Total retail sales of consumer goods	4.4%
Discriminating	Regional total	26.7	Regional	24.6	Added	14.6	Total	7.6%

g	goods amount	%	freight kilometers	%	value of primary industry	%	retail sales of consumer goods	
Cluster 3 Size 4 Isim 0.793 Isdev 0.020 Esim 0.049 Esdev 0.021								
Heze, Zaozhuang, Taian, Liaocheng								
Descriptive	Regional passenger kilometers	11.8 %	The volume of investment in fixed assets	11.4 %	Added value of secondary industry	11.4 %	Population density	10.5 %
Discriminating	Population density	48.5 %	Regional freight kilometers	12.9 %	Regional passenger volume	10.9 %	The volume of investment in fixed assets	7.0%

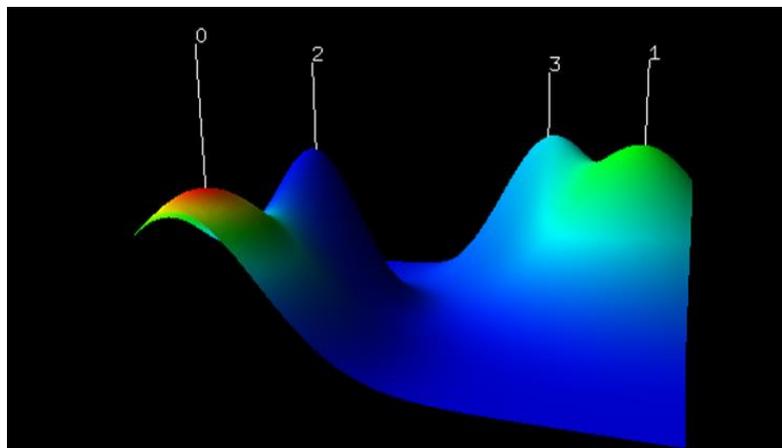
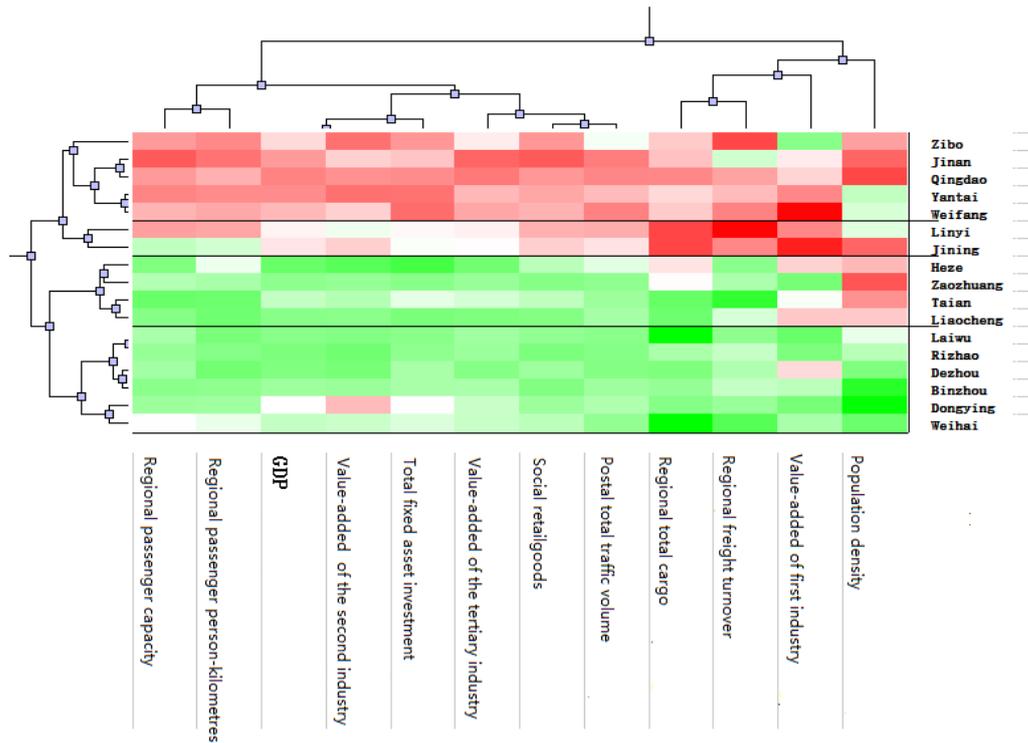


Figure 3-1. Logistics Demand Level Biclustering Visualization Hills



**Figure 3-2. Clustering Results Visualization Matrix of Logistics Demand Levels**

(1) Cluster 0. This cluster contains five cities, Jinan, Qingdao, Zibo, Yantai, Weifang. According to Table 3-1, it can be found that the key indicators are: regional passenger turnover quantity, regional passenger traffic volume, total retail sales of consumer goods and the volume of investment in fixed assets. This means that the matrix of the four cities in these four indicators are a double clustering matrix with high similarity. From the view of quantification, the four indicators are all expressed at high levels. It could be found that the height of this class is lowest from Figure 3-1. It indicates that the similarity in this class is lower than the other three classes. The peak color is red, indicating index distribution is concentrated, small deviations. From the visual matrix (Figure 3-2), it is also noticeable that the five cities are marked with red in most indicators and color differences are small, which means the demand of logistics of the five cities is higher. Class 0 is defined as a high level of the logistics demand of the city group in this article. Remarkably, the five cities have many differences in population density, primary industry and regional freight turnover.

(2) Cluster 1. This class contains six cities, namely Laiwu, Rizhao, Dezhou, Binzhou, Dongying, Weihai. From Table 3-1, it can be found that the key indicators are: population density, regional total cargo, postal total traffic volume and the social retail goods. The internal similarity of this class is the highest among the four classes. According visualization hills (Figure 3-1), it can be found that this class is the highest, indicating the highest internal similarity. The peak is green, which means the distribution is concentrated with small deviation. The visual matrix (Figure 3-2) shows that the values of class 1 in the four indicators are the lowest and the values of other indicators are also lower. Class 1 can be defined as a low level of the logistics demand of the city group.

(3) Cluster 2. This class contains only two cities, Linyi and Jining. From Table 3-1, it can be found that the key indicators are: primary industry, regional freight turnover, regional total cargo and the total retail sales of social consumer goods. According to

visual hills (Figure 3-1), it is remarkable that this class has a higher altitude peaks, indicating higher internal similarity. But the whole mountain is dark blue, which means that the distribution of internal indicators is more dispersed with large deviations. According to visual matrix (Figure 3-2), the values of class 2 in the four indicators are higher but the values of other indicators are almost zero. In addition, Linyi and Jining even show the opposite trend in regional passenger capacity, regional passenger person-kilometers and population density. Even in the presence of expression of the opposite trend. In general, class 2 is a relatively high level of clustering. Therefore, it can be defined as the upper level of the logistics demand of the city group.

(4) Cluster 3. This class contains Heze, Zaozhuang, Tai'an and Liaocheng. From Table 3-1, it can be found that the key indicators are: regional passenger person-kilometers, total fixed asset investment, the secondary industry and population density. From the view of quantification, the value of population density is higher and the other three indicators are lower. Therefore, class 3 can be defined as the lower level of the logistics demand of the city group.

According visual hills (Figure 3-1), it can be found that the height of class 3 is closed to class 1 and class 2, indicating higher internal similarity. The peak is light blue, indicating the distribution of similarity index is scattered with large deviations. From the visual matrix (Figure 3-2), it can be clearly found that the four cities are green on most indicators and with little difference in color, indicating a lower expression values on the corresponding indicators, that means the logistics demand of cities in this class is relatively lower.

### 3.2. Analysis of Logistics Development Levels Based on Biclustering

Within four clusters, ISim are greater than 0.5, ESIm, ISdev, ESdev values are relatively small, indicating good clustering effect. Specific values expression are shown in Table 3-2, 3-3 and Figure 3-4.

**Table 3-2. The City Clustering Result Comparison Table**

Cluster 0 Size 3 Isim 0.892 Isdev 0.024 Esim 0.055 Esdev 0.033										
Laiwu, Zaozhuang, Rizhao										
Descriptive	Highway mileage	36.20%	Freight automobile volume	28.00%	R&D	16.50%	The number of students in College	7.00%	Telephone popularity	5.40%
Discriminating	Highway mileage	60.10%	Freight automobile volume	21.80%	Highway net density	7.40%	R&D	6.40%	Internet popularity	3.70%
Cluster 1 Size 3 Isim 0.808 Isdev 0.058 Esim 0.123 Esdev 0.087										
Weifang, Linyi, Jining										
Descriptive	Highway mileage	37.10%	Freight automobile volume	28.50%	Internet popularity	19.80%	Telephone popularity	10.70%	Highway net density	2.30%
Discriminating	Freight automobile volume	44.50%	Highway mileage	35.50%	Highway net density	9.20%	Internet popularity	4.60%	The number of employee in transportation and inventory	2.40%
Cluster 2 Size 6 Isim 0.653 Isdev 0.162 Esim 0.142 Esdev 0.036										
Zibo, Taian, Bingzhou, Dezhou, Liaocheng, Heze										
Descriptive	Highway net density	44.30%	Internet popularity	12.50%	Telephone popularity	11.40%	The number of employee in transportation and inventory	10.90%	Freight automobile volume	9.20%
Discriminating	Highway net density	69.40%	Freight automobile volume	8.10%	Internet popularity	5.80%	Highway mileage	5.60%	Telephone popularity	5.60%
Cluster 3 Size 5 Isim 0.531 Isdev 0.050 Esim 0.408 Esdev 0.088										
Jinan, Qingdao, Yantai, Weihai, Dongying										
Descriptive	Telephone popularity	24.70%	Internet popularity	23.50%	Highway net density	22.40%	R&D	11.20%	The number of employee in transportation and inventory	7.10%
Discriminating	Telephone popularity	23.40%	Internet popularity	23.20%	Highway net density	19.10%	R&D	11.70%	Internet popularity	9.90%

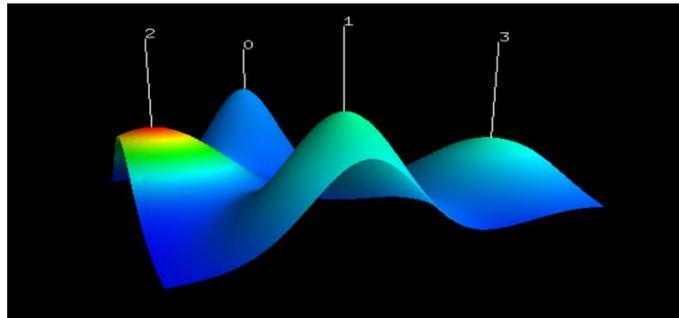


Figure 3-3. Logistics Supply Biclustering Visualization Hills

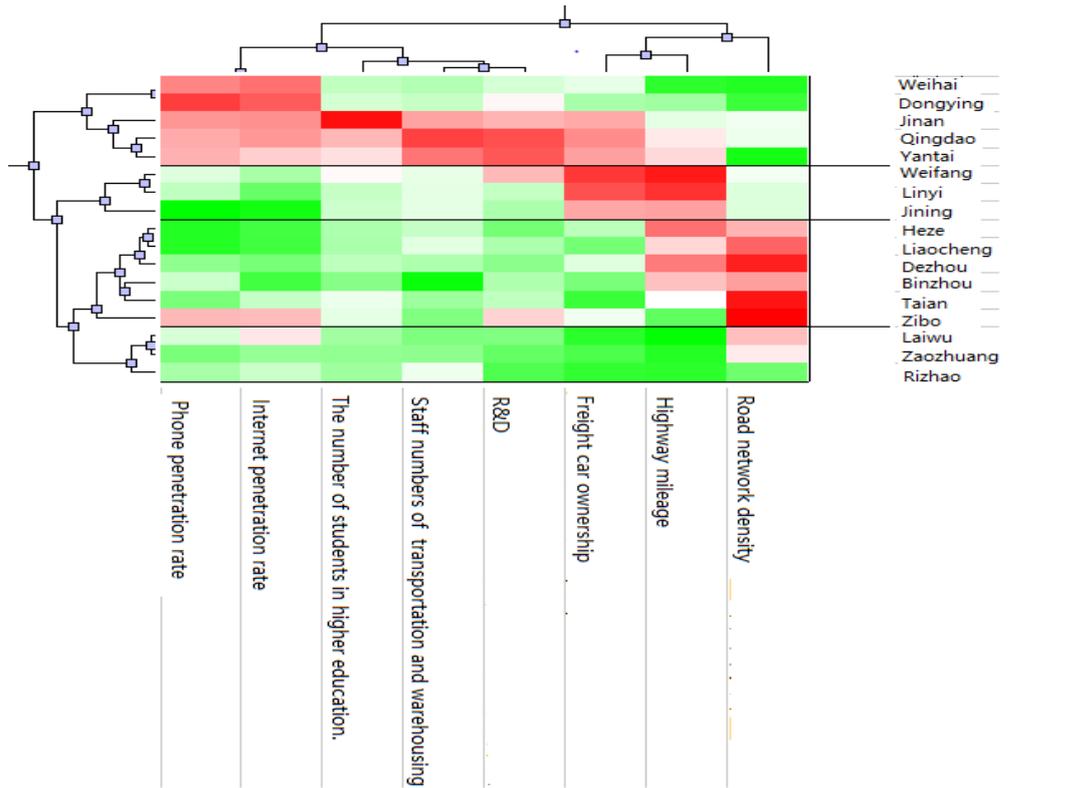


Figure 3-4. Clustering Results Visualization Matrix of Logistics Supply Levels

(1) Cluster 0. This class includes Zaozhuang, Laiwu, Rizhao three cities. We can see from Table 3-2 that describing key indicators of the category such as highway mileage, freight car ownership, the expenditure of big or medium-sized enterprises in research and development test requirements (R&D), and the number of students and telephone penetration in colleges and universities. It indicates that data matrix composed of data node in five indicators in these three cities is a high similarity bi-clustering. According to visual hills (Figure 3-3), we can find that the peak in the category is highest so that it indicates similarity in the category is highest. The peak is blue, which indicates category indicators within categories are scattered distribution and have large deviations. From the visual matrix (Figure 3-4), we can find that in the category the color code of the five cities on most indicators is green, and the color has a small difference, so it indicates that the expression of the corresponding indicators is lowest. In other words, the logistics base supply in the category is low-level. Therefore, Category 0 is defined as low-level logistics supply city group.

(2) Cluster 1. This category only contains Linyi, Jining and Weifang three cities. From Table 3-2, the key describing indicators of the category are: highway mileage, freight car ownership, Internet penetration rate, the telephone penetration rate and density of road network. According to visual hills (Figure 3-3), we can find in this category the altitude is high, so it indicates the similarity is high in the category. The peak color is green, so it indicates that in the category indicators distribution is dispersed, and has large deviations. According to visual matrix (Figure 3-4), the expression of this category in four indicators is high or low. In highway mileage, freight car ownership index, the expression levels of three cities in the province are the highest level, while the three indicators are infrastructure indicators. Road network density of three cities is similar, and they are the middle level in the province. While the telephone penetration and Internet penetration rates are still low. Other indicators of three cities, apart from key describing indicators, are at middle level in the province. On the whole, it is noticeable that cities in this category are at basic level of construction, and have a small number of science and technology talents

(3) Cluster 2. This category concludes Heze, Liaocheng, Dezhou, Binzhou, Tai'an, Zibo. From Table 3-2, we can find that the key describing indicators are the density of the road network, the Internet penetration rate of, telephone penetration, and the number of employees in transportation and warehousing, freight car ownership. According to visual hills (Figure 3-3), we can find that in this category the height of hills is medium, so it indicates the similarity in the category is high; peak color is red, which indicates that the category index distribution is more concentrated, and has small deviations. According to visual matrix (Figure 3-4), we can find the density of road network of the city group is significantly higher than other cities', Internet penetration rate, the telephone penetration rate, and the number of employees in transportation and warehousing, freight car ownership are at a low expression level, which indicates that these urban road constructions are in good condition, but there is obviously insufficient in the cultivation of science and technology talents.

(4) Cluster 3. The category concludes five cities, namely Weihai, Dongying, Jinan, Yantai, Qingdao. In addition to the capital city, Jinan, other cities are economically developed coastal cities. Jinan, Qingdao, as double centers of Shandong Province, have their own characteristics. Jinan is the province's political, economic and cultural center, and is located in the Western. The radiation scope of Jinan includes Dezhou, Liaocheng, Heze, Zibo, Binzhou, Laiwu and so on. Qingdao is an emerging economic and cultural center, promoting the development of Shandong Peninsula city groups such as Yantai, Weihai and Rizhao and so on. As it is shown in Table 3.2, the key describing indicators are: telephone penetration rate, Internet penetration rate, road network density, big and medium-sized industrial enterprises development and testing expenditure as well as number of employees in transportation and warehousing. From the visual matrix (Figure 3-4), we can clearly find in this category the expression values of five cities in terms of telephone and Internet penetration rate are very high, which indicates that the communication levels in these cities are higher. But the road network density is at a low expression level, especially in Weihai, Dongying, Yantai, which are lower than the provincial average. In addition, Jinan, Qingdao, and Yantai is in the forefront of the province in the development and testing expenditure of big and medium-sized industrial enterprises, and the number of employees in transportation and warehousing, while Weihai, Dongying are slightly lower than the provincial average. According to visual hills (Figure 3-3), we can find the height of peak is the lowest, which indicates the similarity in the category is lower. Top color is light green, which indicates similarity index is scattered distributed, and has large deviations. From the visual matrix (Figure 3-4), we can find that the expression level of most of the indicators in this category is in the forefront of the province.

#### 4. Matching Analysis of Logistics Demand and Supply Level

In this paper, we use the panel data of logistics development in 2003-2014 and the double clustering algorithm to cluster the seventeen cities of Shandong Province, and finally divide the seventeen cities into four categories. The final comparison results are shown in Table 4-1.

**Table 4-1. Regional Logistics Demand Scale and Basic Supply Level Comparison Table**

	Demand	supply	
high-level Logistics demand scale	Ji'nan, Qingdao, Yantai, Zibo, Weifang	Ji'nan, Qingdao, Yantai Weihai, Dongying	High level of logistics infrastructure
Middle and upper level Logistics demand scale	Linyi, Jining,	Linyi, Jining, Weifang	High infrastructure construction low science, technology talents and logistics supply ability of the development
Middle and lower level Logistics demand scale	Heze, Tai'an, Liaocheng, Zaozhuang	Heze, Tai'an, Liaocheng, Dezhou, Binzhou, Zibo	High infrastructure construction low science, technology talents and logistics supply ability of the development
low-level Logistics demand scale	Laiwu, Rizhao, Dezhou, Binzhou, Dongying, Weihai	Zaozhuang, Rizhao	Laiwu, Low level of logistics infrastructure

Through comprehensive comparison and analysis, we can found that Jinan, Qingdao and Yantai have high logistics development levels within Shandong province, and that the demand scale and basic supply ability of these cities are relatively high. The equilibrium state of development is a way to achieve the sustainable development of the logistics. These cities should insist strengthening their own development and play the driving role of the central node cities in bid to drive regional development. The urban infrastructure of Linyi, Jining, Heze, Taian and Liaocheng is relatively perfect, however, the cultivation of science and technology talents has been falling behind, which hinders the rapid development of the logistics economy to some extent. Laiwu and Rizhao are in a relatively balanced state of supply and demand, but this is a low level of equilibrium. This backward logistics development status weakened the city's strength. Although most of cities in Shandong province are relatively balanced in the logistics supply and demand, some cities are still harassed by the incoordination of supply and demand.

Mainly manifested in:

(1) Zibo's logistics demands are high, but its basic supply capacity is insufficient, mainly reflected in the low quantity of freight car and the insufficient of cultivating science and technology talents. Similarly, both Weifang and Zibo have the same problems and the density of road network is low.

(2) Weihai and Dongying have higher basic capacity of logistics supply, but the scale of logistics demands is relative small and both cities have excess supply capacity. Weak demand will lead to a waste of resources, and cannot promote healthy competition in the logistics market.

(3) Dezhou and Binzhou have relative small logistics demands. From the point view of supply capacity, the condition of highway infrastructure construction is good. But the lack of logistics talents with scientific and technological skills cannot support the overall

development of the industry in a long term.

(4) In contrast to the situation in Dezhou and Binzhou, Zaozhuang's logistics demands are relatively high, but the basic capacity of supply is seriously insufficient.

## 5. Strategies and Suggestions

As the product of social division of labor, logistics industry has a far-reaching influence on the circulation of the whole social commodity and the allocation of resources, and its importance is self-evident. Accurately positioning the level of urban development, reasonably dividing logistics regions, and correctly understanding the similarities and differences of cities within the same region, can help to better allocate resources and learn from each other in a bid to achieve point to an area strategy and common development.

Based on the results of the analysis of the characteristics of the development of urban logistics, and from the objective needs of the development of the logistics industry, we proposed the following measures to improve the logistics development of Shandong Province.

(1) At present, most of the seventeen cities in the province have the problem of insufficient scientific and technological talents. The measures of the issue include: increasing investment in science and technology, encouraging active and efficient logistics network platform construction, and promoting the integration of business flow, information flow and logistics in a bid to realize the integration and sharing of resources; Besides, it is also essential to build related research institutions, pay closer attention to the trend of the development of the logistics science and technology, boldly adopt innovative technology that suits local development, and encourage the government, enterprise and university researchers explore innovative technology and new standards. Finally, the government should adopt preferential policies to attract logistics professionals, especially senior logistics industry researchers and technical staff.

(2) For cities with the high standard of logistics demands but the low capacity of basic supply, such as Zibo City, can reform investment and financing mechanism of local logistics construction, increasing governments' financial support in the logistics industry and encouraging social investment in in basic logistics industries. We should insist increasing the intensity of the logistics infrastructure, expanding the scope of the transformation, and improving the supply capacity. For cities with the relatively lower standard of logistics demands, such as Dezhou and Binzhou. Government should offer more financial support and policy support, and promote the development of certain leading business in logistics industry. Other measures include promoting the socialization of logistics services, which can be achieved by guiding companies to change their mode of operation and outsourcing inefficient or high cost logistics business to reduce the cost of improving efficiency. It is also important to encourage the corporation between logistics enterprises and circulation enterprises as well as manufacturing enterprises in a bid to enhance the overall competitiveness of the community.

(3) It is wise to break the structure of separately regions in order to build a comprehensive transportation network and an information-sharing platform and promote the standardization of logistics information. The overall utilization efficiency of regional resources can be maxim by using integrated-sharing mode based on the similarity and differences of the urban infrastructure and the characteristics of the resource endowments among cities to achieve the integration and reuse of resources

(4) Sound logistics laws and regulations should be established and preferential policies for the development of logistics should be introduced to constantly improve the environment of logistics development. It is also crucial to optimize industrial structure and encourage logistics industry and related industries to innovate and upgrade.

## 6. Conclusion

The double cluster analysis method has been widely used in the field of biological information, and the row and column are divided to cluster so that the researchers can dig out the similarity between samples from two dimensions. In the analysis of 17 cities in Shandong Province to analyze regional logistics, we use the double clustering analysis method in two dimensions of urban indicators and simultaneously clustering. This can provide not only similar urban agglomeration development status, but also clear similarities and differences of logistics indicators among the cities of the same level. Therefore, local government can introduce policies that suit local development to regulate and support local economy, by the analysis of local geography, economy, culture, transportation and the reference of the development of models in other similar cities.

## Acknowledgments

The first author (Zhang Lei) is partially supported by the Beijing Municipal Commission of Education (No.KM201410016007,21271413117), National Natural Science Foundation of China (No.61402032). The corresponding author (Zhao Peixin) is partially supported by the China Postdoctoral Science Foundation (No.2014T70624) and Natural Science Foundation of Shandong Province (ZR2015GM012).

## References

- [1] J. A. Hartigan, "Direct clustering of a data matrix", *Journal of the American statistical association*, vol. 67, no. 337, (1992), pp. 123-129.
- [2] Y. Cheng and G. M. Church, "Biclustering of expression data", *Proceedings of the Eighth International Conference on Intelligent Systems for Molecular Biology*, (2000), pp. 93-103.
- [3] A. Tanay, R. Sharan, M. Kupiec and R. Shamir, "Revealing modularity and organization in the yeast molecular network by integrated analysis of highly heterogeneous enomewide data", *Proc Natl Acad Sci U S A.*, vol. 101, no. 9, (2004), pp. 2981-6.
- [4] L. Fang and L. Cui, "Research on the subject frontier and intellectual base by using the algorithm: a case study of H index", *Theory & Application*, vol. 11, (2014), pp. 55-60.
- [5] G. Getz, E. Levine and E. Domany, "Coupled two-way clustering analysis of gene microarray data", *Proc. Natl. Acad. Sci. USA*, vol. 97, no. 22, (2000), pp. 12079-84.
- [6] J. Yang, H. Wang, W. Wang and P. Ytr, "Enhanced Bi-clustering Expression Data", *Proc. Third IEEE Conf. Bioinformatics Bioeng.*, (2003), pp. 321-327.
- [7] Y. F. Hou and X. Wang, "The knowledge base and research frontier of detection citation analysis based on Biclustering algorithm", *Journal of Medical Informatics*, vol. 01 (2016), pp. 49-54.
- [8] F. Li, M. Zhang, L. Wang, Y. Ding and L. Cui, "Mining Research Hotspots of International Nursing Informatics by Co-word Analysis", *Journal of Medical Informatics*, vol. 09, (2014), pp. 48-53.
- [9] J. G. Sun, J. Liu and L.Y. Zhao, "Clustering Algorithms Research", *Journal of software*, vol. 01, (2008), pp. 48-61.
- [10] J. Zhao, Q. Ma and Y. Q. Cui, "Comparative study on macro-economic regionalization: an application of biclustering algorithm", *Journal of Shandong University(Natural Science)*, vol. 09, (2012), pp. 71-77.
- [11] D. M. Luo, "Design of the monitoring software for electronic fuel injectors flow characteristics tester based on Labview", *Journal of Luoyang Institute of Science and Technology (Natural Science Edition)*, vol. 04, (2011), pp. 47-49.
- [12] D. Y. Sunaga, J. C. Nievola and M. P. Ramos, "Statistical and Biological Validation Methods in Cluster Analysis of Gene Expression", *International Conference on Machine Learning & Applications*, (2007) pp. 494-499.

## Authors



**Zhang Lei**, lecturer in the Computer Science Department, Beijing University of Civil Engineering and Architecture, Fellow of China Computer Federation(CCF). She received the Ph.D. degree in School of Information and Communication Engineering, Beijing University of Posts and Telecommunications in 2009. Her research interests focus on wireless sensor networks, data security, routing and multicasting protocols and data mining.



**Ma Li**, graduate student of School of Management, Shandong University. Her areas of research are in cluster analysis and logistics management.



**Zhao Peixin**, professor in the School of Management, Shandong University. He received the Ph.D. degree in School of Mathematics, Shandong University in 2005. His research interest focus on data mining and supply chain management.